

Estimator Selection: End-Performance Metric Aspects*

Dimitrios Katselis, Cristian R. Rojas, Carolyn L. Beck

Abstract—Recently, a framework for application-oriented optimal experiment design has been introduced. In this context, the distance of the estimated system from the true one is measured in terms of a particular end-performance metric. This treatment leads to superior unknown system estimates to classical experiment designs based on usual pointwise functional distances of the estimated system from the true one. The separation of the system estimator from the experiment design is done within this new framework by choosing and fixing the estimation method to either a maximum likelihood (ML) approach or a Bayesian estimator such as the minimum mean square error (MMSE). Since the MMSE estimator delivers a system estimate with lower mean square error (MSE) than the ML estimator for finite-length experiments, it is usually considered the best choice in practice in signal processing and control applications. Within the application-oriented framework a related meaningful question is: Are there end-performance metrics for which the ML estimator outperforms the MMSE when the experiment is finite-length? In this paper, we affirmatively answer this question based on a simple linear Gaussian regression example.

I. INTRODUCTION

A basic subproblem in the context of system identification is that of experiment design. Overviews of this topic over the last decade can be found in [5], [7], [15], [8]. Contributions include convexification [10], robust design [13], [16], least-costly design [3], and closed vs open loop experiments [1].

Recently, a new framework for performing experiment design has been introduced. This framework is termed *application-oriented experiment design* and it has been outlined in [8]. Specific investigations related to communication systems were performed in [11], [12]. Denoting the end-performance metric by J and assuming that J depends on the true and the estimated models, the performance is considered to be acceptable if $J \leq 1/\gamma$ for some parameter γ , which we call *accuracy*. This motivates the introduction of a set of admissible models $\mathcal{E}_{adm} = \{G : J \leq 1/\gamma\}$, where G denotes the model to be inferred. With these definitions, the least-costly experiment is formulated as follows:

$$\begin{aligned} & \min_{\text{Experiment}} \quad \text{Experimental effort} \\ & \text{s.t.} \quad \hat{G} \in \mathcal{E}_{adm} \end{aligned} \quad (1)$$

where \hat{G} is the estimated model. For the experimental effort, different measures commonly used are input or output power,

and experimental length. For \hat{G} , standard maximum likelihood (ML) and Bayesian estimation methods, e.g., minimum mean square error (MMSE), are usually employed.

Optimizing the experiment and optimally choosing the system estimator are two problems that should ultimately be tackled in a joint context. Nevertheless, both in the framework of classical and application-oriented experiment designs, a *separation* strategy is applied: initially, we select and fix the system estimator to a choice that is known to possess some optimality aspects, e.g., the ML or MMSE estimators, and then we are optimizing the experiment. For finite-length experiments the MMSE estimator is often considered to be superior to the ML estimator. A related meaningful question in the application-oriented framework is: Are there end-performance metrics for which the ML estimator outperforms the MMSE when the experiment is finite-length?

In this paper, we affirmatively answer the last question based on a simple linear Gaussian regression model that is used here as the simplest possible example to provide the necessary answer. The reason for choosing this example is two-fold: except for the simplicity that it allows, it neutralizes the choice of the optimal experiment. Via this example, we re-examine the validity of the common belief that the MMSE estimator is superior to the ML estimator, when finite length experiments are used to identify the unknown system. To this end, appropriate mean square error (MSE)-like end-performance metrics are used that are meaningful in certain applications such as in communication and control systems. Finally, we numerically demonstrate the validity of the claims verifying the purchased analysis.

This paper is organized as follows: Section II defines the problem of designing the system estimator with respect to the end performance metric. Section III presents some results and comments that will be useful in the rest of the paper, while it introduces approximations of the performance metrics that the rest of the analysis will be based on. The optimality of the ML and MMSE system estimators with respect to the minimization of the aforementioned MSE-like end-performance metrics is examined in Section IV. Section V illustrates the validity of the derived results. Finally, Section VI concludes the paper.

Notations: Vectors are denoted by bold letters. Superscripts T and H stand for transposition and Hermitian transposition, respectively. $|\cdot|$ is the complex modulus. For a vector \mathbf{a} , $a(m)$ denotes its m -th entry. The expectation operator is denoted by $E(\cdot)$. Finally, $\mathcal{CN}(\mu, \sigma^2)$ denotes the complex Gaussian distribution with mean μ and variance σ^2 .

Finalized version.

D. Katselis and C. L. Beck are with the Coordinated Science Laboratory and the Department of Industrial and Enterprise Systems Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801-2925, Emails: {katselis—beck3}@illinois.edu.

Cristian R. Rojas is with the ACCESS Linnaeus Center, Electrical Engineering, KTH Royal Institute of Technology, S-100 44 Stockholm, Sweden. Email: cristian.rojas@ee.kth.se.

II. PROBLEM STATEMENT

Consider the scalar linear Gaussian model

$$y(n) = \theta u(n) + e(n), \quad (2)$$

where $y(n)$ is the observed signal at time instant n , θ is the unknown system parameter assumed to be complex-valued, $u(n)$ is the input at the same time instant and $e(n)$ is complex, circularly symmetric, Gaussian noise with zero mean and variance σ_e^2 . We further assume that $E[u(n)] = 0$ and $E[|u(n)|^2] = \sigma_u^2$. In addition, $w(n)$ and $u(n)$ are independent random sequences, while $e(n)$ is a white random sequence.

Assume that the experimental length is limited to N time slots and that the maximum allowed input energy for experimental purposes is \mathcal{E} . We can collect the received samples corresponding to the experiment in one vector:

$$\mathbf{y}_{\text{exp}} = \theta \mathbf{u}_{\text{exp}} + \mathbf{e}_{\text{exp}}, \quad (3)$$

where $\mathbf{y}_{\text{exp}} = [y(l-N+1), y(l-N+2), \dots, y(l)]^T$ is the vector of N received samples corresponding to the experiment, $\mathbf{u}_{\text{exp}} = [u(l-N+1), u(l-N+2), \dots, u(l)]^T$ is the vector of N input symbols and $\mathbf{e}_{\text{exp}} = [e(l-N+1), e(l-N+2), \dots, e(l)]^T$ is the vector of N noise samples. Considering the class of linear parameter estimators, the system is estimated as follows:

$$\hat{\theta} = \mathbf{f}^H \mathbf{y}_{\text{exp}} = \theta \mathbf{f}^H \mathbf{u}_{\text{exp}} + \mathbf{f}^H \mathbf{e}_{\text{exp}}, \quad (4)$$

where \mathbf{f} is a $N \times 1$ estimating filter.

A possible performance metric is the MSE of a *linear* input estimator. The input estimator uses the system knowledge and delivers an estimate of the input variable. We call *clairvoyant* the input estimator that has perfect system knowledge. Denoting the corresponding estimating filter by $\tilde{c}(\theta)$, we can find its mathematical expression as follows:

$$\tilde{c}(\theta) = \arg \min_{c(\theta)} E \left[|c(\theta)y(n) - u(n)|^2 \right], \quad (5)$$

where the expectation is taken over the statistics of $u(n)$ and $e(n)$. If we set the derivative of the last expression with respect to $c(\theta)$ to zero and we solve for $c(\theta)$, then the optimal clairvoyant input estimating filter is given by the expression

$$\tilde{c}(\theta) = \frac{\sigma_u^2 \theta^*}{|\theta|^2 \sigma_u^2 + \sigma_e^2}. \quad (6)$$

We will call this the MMSE clairvoyant input estimator¹. We observe that as the signal-to-noise ratio (SNR) increases, i.e., $\sigma_e^2 \rightarrow 0$, $\tilde{c}(\theta) \rightarrow 1/\theta$. We call $\tilde{c}(\theta) = 1/\theta$ the *Zero Forcing* (ZF) clairvoyant input estimator. Due to this last convergence and for simplicity purposes, we focus only on the ZF input estimator in the sequel.

We can now introduce an end-performance metric of interest, which will be used in the following analysis. Given an input estimator, we define the excess of the input estimate

based on an input estimator that only knows a system estimate over the input estimator with perfect system knowledge, thus leading to

$$\text{MSE}_{ex} = E \left[|c(\hat{\theta})y(n) - c(\theta)y(n)|^2 \right]. \quad (7)$$

In the sequel, this metric will be called *excess* MSE.

Our goal will be to determine the optimal parameter estimators for fixed experiments of finite length so that MSE_{ex} based on the ZF input estimator is minimized. To this end, the following section presents some useful ideas.

III. PRELIMINARY RESULTS

Consider the ML estimator. For the linear Gaussian regression, this estimator coincides with the minimum variance unbiased (MVU) estimator. We therefore replace our references to the ML estimator by references to the MVU estimator from now on. Since the MVU is an unbiased estimator, it satisfies $\mathbf{f}^H \mathbf{u}_{\text{exp}} = 1$. This condition implies that $E[\hat{\theta}] = \theta$. For our problem assumptions, the MVU estimator can be found by solving the following optimization problem:

$$\begin{aligned} \min_{\mathbf{f}} \quad & \sigma_e^2 \|\mathbf{f}\|^2 \\ \text{s.t.} \quad & \mathbf{f}^H \mathbf{u}_{\text{exp}} = 1. \end{aligned} \quad (8)$$

Forming the Lagrangian for this problem and zeroing its gradient with respect to \mathbf{f} , we get:

$$\mathbf{f}_{\text{MVU}} = \frac{\mathbf{u}_{\text{exp}}}{\|\mathbf{u}_{\text{exp}}\|^2}. \quad (9)$$

If we assume that θ is a random variable and that its prior distribution is known, then instead of the MVU one could use the MMSE parameter estimator. With our assumptions and the extra assumption that $E[\theta] = 0$, one can obtain [14]

$$\mathbf{f}_{\text{MMSE}} = \frac{E[|\theta|^2] \mathbf{u}_{\text{exp}}}{E[|\theta|^2] \|\mathbf{u}_{\text{exp}}\|^2 + \sigma_e^2}. \quad (10)$$

Assuming that θ is a deterministic but unknown variable, the MSE_{ex} of the ZF input estimator can be easily obtained:

$$\text{MSE}_{ex}^d(\text{ZF}) = E \left[\left| \frac{\hat{\theta} - \theta}{\hat{\theta}} \right|^2 \right] \left(\sigma_u^2 + \frac{\sigma_e^2}{|\theta|^2} \right) \quad (11)$$

(c.f. (7)). Here, the superscript “d” stands for “deterministic”. If θ is assumed to be a random variable, then the corresponding end-performance metric MSE_{ex}^r is obtained by averaging the last expression over θ .

Depending on the probability distributions of $|\hat{\theta}|$ and $|\theta|$, the above MSE expressions may fail to exist. The MSEs will be finite if the probability distribution function (pdf) of $|\hat{\theta}|$ is of order $O(|\hat{\theta}|^2)$ as $\hat{\theta} \rightarrow 0$. A similar condition should hold for the pdf of $|\theta|$ in the case of MSE_{ex}^r . In the opposite case, we end up with an *infinite moment* problem. In order to obtain well-behaved parameter estimators that will be used in conjunction with the actual performance metric, some sort of regularization is needed. Some ideas for appropriate regularization techniques to use may be

¹The multiplication by $y(n)$ is considered implicit.

obtained by modifying robust estimators (against heavy-tailed distributions), e.g., by trimming a standard estimator, if it gives a value very close to zero [9]. An example of such a trimmed estimator is given as follows:

$$\hat{\theta} = \begin{cases} \mathbf{f}^H \mathbf{y}_{\text{exp}}, & \text{if } |\mathbf{f}^H \mathbf{y}_{\text{exp}}| > \lambda \\ \lambda \mathbf{f}^H \mathbf{y}_{\text{exp}} / |\mathbf{f}^H \mathbf{y}_{\text{exp}}|, & \text{o.w.} \end{cases} \quad (12)$$

where \mathbf{f} can be any estimator and λ a regularization parameter².

Remark: Clearly, the reader may observe that the definition of the trimmed $\hat{\theta}$ preserves the continuity at $|\mathbf{f}^H \mathbf{y}_{\text{exp}}| = \lambda$. Additionally, the event $\{\mathbf{f}^H \mathbf{y}_{\text{exp}} = 0\}$ has zero probability since the distribution of $\mathbf{f}^H \mathbf{y}_{\text{exp}}$ is continuous. Therefore, in this case $\hat{\theta}$ can be arbitrarily defined, e.g., $\hat{\theta} = \lambda$.

Assume a fixed λ . Then, for a sufficiently small λ and a sufficiently high SNR during training, minimizing $\text{MSE}_{ex}^d(\text{ZF})$ is approximately equivalent to minimizing the approximation

$$\left[\text{MSE}_{ex}^d(\text{ZF})\right]_0 = \frac{E\left[\left|\hat{\theta} - \theta\right|^2\right]}{E\left[\left|\hat{\theta}\right|^2\right]} \left(\sigma_u^2 + \frac{\sigma_e^2}{|\theta|^2}\right), \quad (13)$$

as we show in the appendix. Using some minor additional technicalities, we can work with

$$\left[\text{MSE}_{ex}^r(\text{ZF})\right]_0 = \frac{\sigma_u^2 E_\theta \left[|\theta|^2 E\left[\left|\hat{\theta} - \theta\right|^2\right]\right] + \sigma_e^2 E_\theta \left[E\left[\left|\hat{\theta} - \theta\right|^2\right]\right]}{E_\theta \left[|\theta|^2 E\left[\left|\hat{\theta}\right|^2\right]\right]}, \quad (14)$$

instead of $\text{MSE}_{ex}^r(\text{ZF})$. We call the last approximations *zeroth order* input estimate excess MSEs. The following analysis and results will be based on the zeroth order metrics and they will reveal the dependency of the system estimator's selection on the considered (any) end-performance metric.

Remarks:

- 1) A useful, alternative way to consider the zeroth order MSEs is to view them as affine versions of normalized parameter MSEs, where the actual true parameter is $\hat{\theta}$ and the estimator is θ .
- 2) In the definition of (13), one can observe that after approximating the mean value of the ratio by the ratio of the mean values the infinite moment problem is eliminated. In the following, all zeroth order metrics will be defined based on the *non-trimmed* $\hat{\theta}$ to ease the derivations. This treatment is approximately valid when λ is sufficiently small.

IV. MINIMIZING THE ZEROth ORDER EXCESS MSE

In this section, we investigate the selection of the system estimator for the zeroth order excess MSE in the case of the ZF input estimator.

²This parameter can be tuned via cross-validation or any other technique, although in the simulation section we empirically select it for simplicity purposes.

A. ZF Input Estimator with a Deterministic System

The expectation operators in Eq. (13) are with respect to $\mathbf{e}_{\text{exp}}, u(n)$ and $e(n)$. In this case, we have:

$$\left[\text{MSE}_{ex}^d(\text{ZF})\right]_0 = \frac{|\theta|^2 |\mathbf{f}^H \mathbf{u}_{\text{exp}} - 1|^2 + \sigma_e^2 \|\mathbf{f}\|^2}{|\theta|^2 |\mathbf{f}^H \mathbf{u}_{\text{exp}}|^2 + \sigma_e^2 \|\mathbf{f}\|^2} \left(\sigma_u^2 + \frac{\sigma_e^2}{|\theta|^2}\right) \quad (15)$$

The numerator of the gradient of the above expression with respect to³ \mathbf{f} is given by the following expression:

$$\begin{aligned} & [|\theta|^2 |\varphi|^2 + \sigma_e^2 \|\mathbf{f}\|^2] [|\theta|^2 (\varphi - 1)^* \mathbf{u}_{\text{exp}} + \sigma_e^2 \mathbf{f}] \\ & - [|\theta|^2 \varphi^* \mathbf{u}_{\text{exp}} + \sigma_e^2 \mathbf{f}] [|\theta|^2 |\varphi - 1|^2 + \sigma_e^2 \|\mathbf{f}\|^2], \end{aligned} \quad (16)$$

where $\varphi = \mathbf{f}^H \mathbf{u}_{\text{exp}}$. Setting $\mathbf{f} = \mathbf{f}_{\text{MVU}}$, one can easily check that the above expression becomes zero. Therefore:

Proposition 1: The MVU is an optimal system estimator for the task of minimizing $\left[\text{MSE}_{ex}^d(\text{ZF})\right]_0$, when the system parameter is considered a deterministic but otherwise unknown quantity.

Remark: Note that even if $\left[\text{MSE}_{ex}^d(\text{ZF})\right]_0$ depends on the unknown system parameter θ , the optimal system estimator does not in this case.

B. ZF Input Estimator with a Random System

In this case, the prior statistics of θ are known. The zeroth order excess MSE is given by:

$$\begin{aligned} \left[\text{MSE}_{ex}^r(\text{ZF})\right]_0 &= \frac{|\varphi - 1|^2 (E[|\theta|^4] \sigma_u^2 + E[|\theta|^2] \sigma_e^2)}{E[|\theta|^4] |\varphi|^2 + \sigma_e^2 \|\mathbf{f}\|^2 E[|\theta|^2]} \\ &+ \frac{\sigma_e^2 \|\mathbf{f}\|^2 (E[|\theta|^2] \sigma_u^2 + \sigma_e^2)}{E[|\theta|^4] |\varphi|^2 + \sigma_e^2 \|\mathbf{f}\|^2 E[|\theta|^2]} \end{aligned} \quad (17)$$

Differentiating this expression w.r.t. \mathbf{f} and setting $\mathbf{f} = \mathbf{f}_{\text{MVU}}$ we zero the gradient. Therefore:

Proposition 2: The MVU is an optimal system estimator for the task of minimizing $\left[\text{MSE}_{ex}^r(\text{ZF})\right]_0$, when the system parameter is considered random.

Via tedious calculations, we can show that the MMSE channel estimator does not zero the gradient.

Remark: This result is *counterintuitive*: it says that when one has knowledge of the system statistics but uses a ZF input estimator, one should ignore these statistics in choosing a system estimator for minimizing the zeroth order excess MSE. This is the *major* result in this paper: The belief that combining the MMSE system estimator with any performance metric is better than using the MVU/ML system estimator when finite length experiments are used to identify the system, is *not* valid.

C. Discussion on the Optimal Training

Since the system estimator is selected in order to optimize the final performance metric, one may consider the problem of selecting optimally the input vector \mathbf{u}_{exp} under a maximum energy constraint $\|\mathbf{u}_{\text{exp}}\|^2 \leq \mathcal{E}$ to serve the same

³discarding the positive scalars and considering again the corresponding (hermitian) transpositions.

purpose. To optimize the input vector, one should first fix the system estimator. This is a “complementary” problem with respect to the approach that we have followed so far. Suppose that we use either the MVU or the MMSE system estimators. One can observe that for $N = 1$ the problem of selecting optimally the input vector is meaningless. In the case that $N > 1$, fixing for example $\mathbf{f} = \mathbf{f}_{\text{MVU}}$ one can observe that again the problem of selecting optimally the input vector is meaningless. Consider for example the case of $[\text{MSE}_{ex}^r(\text{ZF})]_0$. We then have:

$$[\text{MSE}_{ex}^r(\text{ZF})]_0 = \frac{\sigma_e^2 (E[|\theta|^2] \sigma_u^2 + \sigma_e^2)}{E[|\theta|^4] \|\mathbf{u}_{\text{exp}}\|^2 + \sigma_e^2 E[|\theta|^2]},$$

which only depends on $\|\mathbf{u}_{\text{exp}}\|^2$. Furthermore, $[\text{MSE}_{ex}^r(\text{ZF})]_0$ is minimized when $\|\mathbf{u}_{\text{exp}}\|^2 = \mathcal{E}$, which is intuitively appealing. Therefore, any \mathbf{u}_{exp} with energy equal to \mathcal{E} is an equally good input vector for the MVU estimator. Thus, for the same \mathbf{u}_{exp} , the MVU estimator is better than the MMSE.

V. SIMULATIONS

In this section we present numerical results to verify our analysis. In all figures, $\theta \sim \mathcal{CN}(0, 1)$. The SNR during the experiment highlights how good the system estimate is. The parameter λ has been empirically selected to be 0.1 in Fig. 2. The two figures that we present in this section aim at two goals: first, to highlight that indeed the MVU/ML estimator can be better than the MMSE in finite length system identification depending on the end-performance metric of interest. And second, to verify that the zeroth order approximations used in this paper for analysis purposes are good approximations to the true end-performance metrics for extracting the necessary conclusions.

Fig. 1 presents the corresponding results for $[\text{MSE}_{ex}^r(\text{ZF})]_0$. The SNR during the experiment has been set to 0 dB, which can be a low operational value in real world applications, but useful, e.g., in situations where energy efficiency is crucial such as in wireless sensor networks. The experimental length has been set to 2 simply to eliminate the asymptotic efficiency of the ML estimator. The MVU is the best estimator as proven. This is an example contradicting what one would expect and verifying the motivation of this paper.

Finally, Fig. 2 verifies that the zeroth order metrics used in this paper are good approximations in terms of indicating the structure of uniformly better estimators than the MMSE. The SNR during the experiment and the experimental length are as before. We observe that except for a translation in the vertical direction, the zeroth order approximations are able to indicate the relative position of the estimating curves leading to accurate conclusions about the comparison between them.

VI. CONCLUSIONS

In this paper, end-performance metric system estimator selection has been investigated. We have shown that the application-oriented selection is the right way to choose estimators in practice. We have verified this observation based on

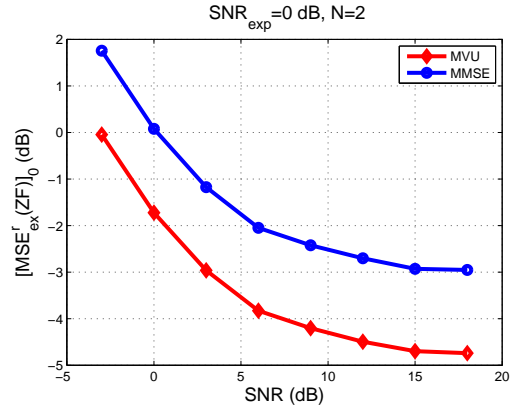


Fig. 1. $[\text{MSE}_{ex}^r(\text{ZF})]_0$ with SNR during the experiment equal to 0 dB and $N = 2$.

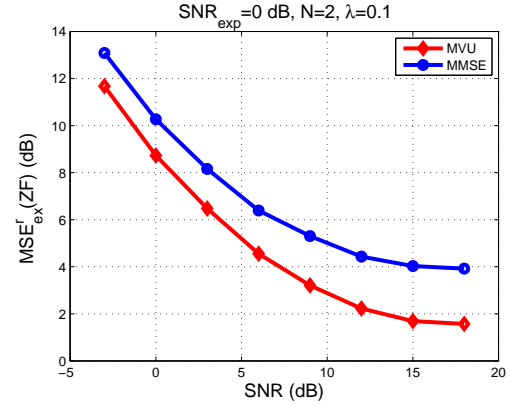


Fig. 2. $\text{MSE}_{ex}^r(\text{ZF})$ with SNR during the experiment equal to 0 dB, $N = 2$ and $\lambda = 0.1$.

an explanatory end-performance metric of interest, namely, the excess input estimate MSE. The extracted conclusion is that the ML/MVU estimators can be better than the MMSE estimator for particular end-performance metrics of interest. This invalidates the common belief that the MMSE estimator is always better than the ML/MVU estimators for *any* end-performance metric, if finite length experiments are used for system identification purposes.

APPENDIX

This section proposes a simplification of the $[\text{MSE}_{ex}^d(\text{ZF})]$ metric for the estimator given in (12) with a fixed λ . Due to the Gaussianity of \mathbf{u}_{exp} , $[\text{MSE}_{ex}^d(\text{ZF})] = \infty$ for any $\mathbf{f} \neq \mathbf{0}$ (infinite moment

problem). Using (12), the corresponding metric becomes:

$$\begin{aligned} [\text{MSE}_{ex}^d(\text{ZF})]_{\text{reg}} &= \Pr \left\{ |\mathbf{f}^H \mathbf{y}_{\text{exp}}| > \lambda \right\} \cdot \\ &E \left[\left(\sigma_u^2 + \frac{\sigma_e^2}{|\theta|^2} \right) \left| 1 - \frac{\theta}{\mathbf{f}^H \mathbf{y}_{\text{exp}}} \right|^2 ; |\mathbf{f}^H \mathbf{y}_{\text{exp}}| > \lambda \right] \\ &+ \Pr \left\{ |\mathbf{f}^H \mathbf{y}_{\text{exp}}| \leq \lambda \right\} \cdot \\ &E \left[\left(\frac{\sigma_u^2}{\lambda^2} + \frac{\sigma_e^2}{\lambda^2 |\theta|^2} \right) \left| \lambda \frac{\mathbf{f}^H \mathbf{y}_{\text{exp}}}{|\mathbf{f}^H \mathbf{y}_{\text{exp}}|} - \theta \right|^2 ; |\mathbf{f}^H \mathbf{y}_{\text{exp}}| \leq \lambda \right], \end{aligned}$$

where ; denotes conditioning and “reg” signifies the use of the regularized system estimator in (12). Moreover, $\Pr \left\{ |\mathbf{f}^H \mathbf{y}_{\text{exp}}| \leq \lambda \right\} = O(\lambda^2)$, since by the mean value theorem this probability is equal to the area of the region $\{|\mathbf{f}^H \mathbf{y}_{\text{exp}}| \leq \lambda\}$, which is of order $O(\lambda^2)$, multiplied by some value of the probability density function of $|\mathbf{f}^H \mathbf{y}_{\text{exp}}|$ in that region, which is of order $O(1)$. In addition,

$$\begin{aligned} &E \left[\left(\frac{\sigma_u^2}{\lambda^2} + \frac{\sigma_e^2}{\lambda^2 |\theta|^2} \right) \left| \lambda \frac{\mathbf{f}^H \mathbf{y}_{\text{exp}}}{|\mathbf{f}^H \mathbf{y}_{\text{exp}}|} - \theta \right|^2 ; |\mathbf{f}^H \mathbf{y}_{\text{exp}}| \leq \lambda \right] \\ &= \left(\sigma_u^2 + \frac{\sigma_e^2}{|\theta|^2} \right) + \left(\frac{\sigma_u^2}{\lambda^2} |\theta|^2 + \frac{\sigma_e^2}{\lambda^2} \right) \\ &- 2 \left(\frac{\sigma_u^2}{\lambda} + \frac{\sigma_e^2}{\lambda |\theta|^2} \right) E \left[\Re \left\{ \theta^* \frac{\mathbf{f}^H \mathbf{y}_{\text{exp}}}{|\mathbf{f}^H \mathbf{y}_{\text{exp}}|} \right\} \right]. \end{aligned}$$

Furthermore, if the SNR during training is sufficiently high and the probability mass of $|\mathbf{f}^H \mathbf{y}_{\text{exp}}|$ is concentrated around $|\theta|$, then it can be shown that

$$\begin{aligned} &E \left[\left(\sigma_u^2 + \frac{\sigma_e^2}{|\theta|^2} \right) \left| 1 - \frac{\theta}{\mathbf{f}^H \mathbf{y}_{\text{exp}}} \right|^2 ; |\mathbf{f}^H \mathbf{y}_{\text{exp}}| > \lambda \right] \\ &\approx \frac{(\sigma_u^2 + \sigma_e^2/|\theta|^2) E[|\mathbf{f}^H \mathbf{y}_{\text{exp}} - \theta|^2 ; |\mathbf{f}^H \mathbf{y}_{\text{exp}}| > \lambda]}{E[|\mathbf{f}^H \mathbf{y}_{\text{exp}}|^2 ; |\mathbf{f}^H \mathbf{y}_{\text{exp}}| > \lambda]}. \end{aligned} \quad (19)$$

The same holds even if $\mathbf{f}^H \mathbf{y}_{\text{exp}}$ is a biased estimator of θ at high training SNR and $|\mathbf{f}^H \mathbf{y}_{\text{exp}}|$ tends to concentrate around a value β bounded away from $|\theta|$ (and of course from 0).

To show the last claim, we set $X = |\mathbf{f}^H \mathbf{y}_{\text{exp}} - \theta|^2$ and $Y = |\mathbf{f}^H \mathbf{y}_{\text{exp}}|^2$. Since $Y > \lambda^2$, it also holds that $E[Y] > \lambda^2$. Furthermore, it can be seen that

$$\left| E \left[\frac{X}{Y} \right] - \frac{E[X]}{E[Y]} \right| \leq \frac{1}{\lambda^4} E[|XE[Y] - YE[X]|]. \quad (20)$$

At high training SNR, $X \rightarrow E[X]$ and $Y \rightarrow E[Y]$ in the mean square sense and therefore it can be easily shown that the right hand side of (20) converges to 0. To see this, notice that the Cauchy-Schwarz inequality yields

$$\begin{aligned} &\frac{1}{\lambda^4} E[|XE[Y] - YE[X]|] \leq \frac{1}{\lambda^4} \left(E[|XE[Y] - YE[X]|^2] \right)^{1/2} \\ &= \frac{1}{\lambda^4} (E^2[Y]E[X^2] + E[Y^2]E^2[X] - 2E[XY]E[X]E[Y])^{1/2}. \end{aligned} \quad (21)$$

Since $X \rightarrow E[X]$ and $Y \rightarrow E[Y]$ in the mean square sense, $E[X^2] \rightarrow E^2[X]$, $E[Y^2] \rightarrow E^2[Y]$ and $E[XY] \rightarrow E[X]E[Y]$. For the last case, notice that

$$|E[XY] - E[X]E[Y]| \leq \sqrt{E[|X - E[X]|^2] E[|Y - E[Y]|^2]},$$

where the last inequality follows again from the Cauchy-Schwarz inequality. By the mean square convergence of X to $E[X]$ and Y to $E[Y]$ the right hand side of the last inequality tends to 0. Therefore, the right hand side of (21) tends to 0.

Moreover, under the high SNR assumption the conditional (18) expectations can be approximated by their unconditional ones, since for a sufficiently small λ their difference is due to an event of probability $O(\lambda^2)$. Therefore,

$$\begin{aligned} &E \left[\left(\sigma_u^2 + \frac{\sigma_e^2}{|\theta|^2} \right) \left| 1 - \frac{\theta}{\mathbf{f}^H \mathbf{y}_{\text{exp}}} \right|^2 ; |\mathbf{f}^H \mathbf{y}_{\text{exp}}| > \lambda \right] \approx \\ &\left\{ \frac{(\sigma_u^2 + \sigma_e^2/|\theta|^2) E[|\mathbf{f}^H \mathbf{y}_{\text{exp}} - \theta|^2]}{E[|\mathbf{f}^H \mathbf{y}_{\text{exp}}|^2]} \right\} + O(\lambda^2). \end{aligned} \quad (22)$$

Combining all the above results yields

$$[\text{MSE}_{ex}^d(\text{ZF})]_{\text{reg}} \approx \left\{ \frac{(\sigma_u^2 + \sigma_e^2/|\theta|^2) E[|\mathbf{f}^H \mathbf{y}_{\text{exp}} - \theta|^2]}{E[|\mathbf{f}^H \mathbf{y}_{\text{exp}}|^2]} \right\} + O(1). \quad (23)$$

The $O(1)$ term is not negligible but for sufficiently small λ its dependence on \mathbf{f} is insignificant. Hence, for a sufficiently small λ and a sufficiently high SNR during training, minimizing $[\text{MSE}_{ex}^d(\text{ZF})]_{\text{reg}}$ is equivalent to minimizing (11).

REFERENCES

- [1] J. C. Aguero and G. C. Goodwin, “Choosing between open and closed-loop experiments in linear system identification”, *IEEE Transactions on Automatic Control*, vol. 52, no. 8, pp. 1475–1480, Aug. 2007.
- [2] M. Barenthin, X. Bombois, H. Hjalmarsson, and G. Scorletti, “Identification for control of multivariable systems: Controller validation and experiment design via lmis”, *Automatica*, vol. 44, no. 12, pp. 3070–3078, Dec. 2008.
- [3] X. Bombois, G. Scorletti, M. Gevers, P. Van den Hof, and R. Hildebrand, “Least costly identification experiment for control”, *Automatica*, vol. 42, no. 10, pp. 1651–1662, Oct. 2006.
- [4] U. Forssell and L. Ljung, “Some results on optimal experiment design”, *Automatica*, vol. 36, no. 5, pp. 749–756, May 2000.
- [5] M. Gevers, “Identification for control: From the early achievements to the revival of experiment design”, *European Journal of Control*, 11:1–18, 2005.
- [6] M. Gevers and L. Ljung, “Optimal experiment designs with respect to the intended model application”, *Automatica*, vol. 22, no. 5, pp. 543–554, 1986.
- [7] H. Hjalmarsson, “From experiment design to closed-loop control”, *Automatica*, vol. 41, no. 3, pp. 393–438, March 2005.
- [8] H. Hjalmarsson, “System identification of complex and structured systems”, *Plenary address European Control Conference / European Journal of Control*, vol. 15, no. 4, pp. 275–310, 2009.
- [9] P. J. Huber, *Robust Statistics*, John Wiley & Sons, 2005.
- [10] H. Jansson and H. Hjalmarsson, “Input design via LMIs admitting frequency-wise model specifications in confidence regions”, *IEEE Trans. on Automatic Control*, vol. 50, no. 10, pp. 1534–1549, Oct. 2005.
- [11] D. Katselis, C. R. Rojas, H. Hjalmarsson, and M. Bengtsson, “Application-oriented finite sample experiment design: A semidefinite relaxation approach”, *Proc. SYSID 2012*, Brussels, Belgium, July 2012.

- [12] D. Katselis, C. R. Rojas, M. Bengtsson, E. Bjrnson, X. Bombois, N. Shariati, M. Jansson, H. Hjalmarsson, "Training sequence design for MIMO channels: an application-oriented approach", . *EURASIP J. Wireless Comm. and Networking* 2013, 2013: 245.
- [13] D. Katselis, C. R. Rojas, J. S. Welsh, H. Hjalmarsson, " Robust Experiment Design for System Identification via Semi-Infinite Programming Techniques", *Proc. SYSID 2012*, Brussels, Belgium, July 2012.
- [14] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, Prentice Hall, 1993.
- [15] L. Pronzato, "Optimal experimental design and some related control problems", *Automatica*, vol. 44, no. 2, pp. 303–325, Feb. 2008.
- [16] C. R. Rojas, J. S. Welsh, G. C. Goodwin, and A. Feuer, "Robust optimal experiment design for system identification", *Automatica*, vol. 43, no. 6, pp. 993–1008, 2007.